Sites Inventory Appendix B2:
Modeling Development on Non-Vacant and Vacant Sites

ADOPTED - January 2023

Prepared as Part of the
San Francisco Planning Department's

Housing Element Update 2022

San Francisco Planning

Technical Appendix: Modeling Development on Non-Vacant and Vacant Sites

Contents

Introduction

Much of the development expected to occur in San Francisco over the coming 8 years is anticipated as part of the residential development pipeline as well as in projects that are subject to development agreements and a smaller number of units to be built as ADUs[1]. This appendix describes the methodology employed by San Francisco to estimate the likely extent of development on the remaining sites in the City.

Changes to State Housing Element law (particularly Assembly Bill 1397 passed in 2017) have strengthened requirements that sites included in the inventory be realistically assessed for their development potential within the 8-year RHNA planning period. When the sites inventory includes more than 50% non-vacant sites, existing uses are presumed to impede development unless substantial evidence is provided that the use is likely to be discontinued. In San Francisco nearly all land available for residential development is not vacant and the approach to assessing development potential to accommodate RHNA must realistically address this fact.[2] While San Francisco has ample examples of non-vacant sites redeveloping as housing, the methodology used to identify realistic development potential must consider factors such as existing uses, past development trends, market conditions, and other factors relevant to whether sites can realistically be redeveloped as housing.

The approach employed by San Francisco involves analyzing past development activity (including on non-vacant sites) and identifying the parcel characteristics and housing market and economic conditions that prevailed at the time of development in order to develop a model to estimate the likelihood of development on individual sites going forward. Specifically, in order to estimate the impact of housing policies and market conditions on the extent and location of new housing development in San Francisco, the Planning Department contracted with the Blue Sky Consulting Group to conduct an analysis of San Francisco housing development trends as part of the Housing Affordability Strategies (HAS) project completed in 2020. This analysis was updated in 2022 for purposes of using these results in the preparation of this report. The housing market analysis was conducted using a logistic regression in which the likelihood of market-rate multifamily housing development was estimated based on a series of explanatory variables, including construction costs, housing prices, and parcel-specific characteristics including contemporaneous zoning category, current residential use or historical designation, current permissible building size (envelope), and development potential (ratio of permissible to existing building size). Results of the regression analysis are presented in Figure 4 on page 10, which shows that each of the key explanatory variables was highly statistically significant.

Large project areas, such as Treasure Island or Mission Bay, were estimated separately by Planning in collaboration with the Office of Housing Delivery, other City agencies, and developers based on the specifics of the development agreements covering these projects. Projects already in the development

1 Note: the ADU estimates exclude any impacts stemming from SB 9, which are modeled as part of underlying analysis.

2 See Appendix 3 for a series of case studies of sites that developed as housing.

pipeline, (non-inclusionary) affordable units, and accessory dwelling units were also estimated by Planning separately.[3,4]

Period of Study

The Blue Sky Consulting Group analyzed housing development during the period 2001-2018. Data for the period 2019-2021 were incorporated into the analysis for purposes of identifying new residential development that occurred during this period, capturing any changes to zoning or parcel characteristics, and incorporating current construction cost and price data to reflect current economic conditions driving housing production. The underlying statistical relationships used to derive the model results were not re-estimated due to the likely confounding effects of the COVID 19 pandemic.

Current Zoned Capacity and Historical Development Activity

San Francisco's current zoned capacity could more than accommodate the entire 8-year RHNA target (if all sites developed). As shown in Figure 1, the zoned capacity on sites covered by this analysis (i.e. any site not part of a development agreement, in the current project pipeline, or otherwise excluded due to the low likelihood of future development of housing, such as historical sites) is almost 640,000 units if the maximum state density bonus is applied to all eligible parcels, and over 570,000 units when the bonus is applied consistent with historical patterns.[5]

Figure 1. Zoned Capacity

	< 10 Units	*%*	*10 – 50 Units*	*%*	*> 50 Units*	*%*	*Total*
At Maximum Development Potential:							
Parcels	141,033	95.4%	5,069	3.4%	1,686	1.1%	**147,788**
Net Units	288,076	45.1%	82,983	13.0%	268,061	41.9%	**639,120**
At Modeled Development Potential:							
Parcels	141,245	95.6%	5,135	3.5%	1,408	1.0%	**147,788**
Net Units	289,166	50.5%	74,191	13.0%	209,078	36.5%	**572,434**

Notes: Data include only parcels used in the estimation model.

Maximum Development Potential is estimated using 50%state density bonus for all eligible parcels, while the model relies on historical patterns to apply a 40% density bonus to 60% of eligible parcels.

3 Development of subsidized affordable housing was analyzed separately but was not included in the model developed by the Blue Sky Consulting Group as the characteristics of these projects and the market conditions that can make them feasible are distinct in many respects from the factors that drive market rate or privately financed development.

4 Parcels in the development pipeline with a non-residential planned use were also excluded from the model as these parcels are unlikely to be a source of future housing development during the RHNA period.

5 Over the past several years, approximately 60 percent of multi-family projects have used the state density bonus.

Zoned capacity, however, will not necessarily translate into actual housing units to the extent market conditions make development infeasible or regulatory barriers or other housing policies prevent development from occurring. During the period 2001 – 2022, for example, approximately 2,700 housing units were added annually in San Francisco, as shown in Figure 2. The average over the past 10 years was somewhat higher, with nearly 3,500 units added annually.

Figure 2. San Francisco Housing Trends

Year	*Net Units Authorized*	*Units Completed from New Construction*	*Units Demolished*	*Net Units Gained or Lost from Alterations*	*Net Change in Number of Units*
2001	2,380	1,619	99	259	1,779
2002	1,478	2,260	73	221	2,408
2003	1,845	2,730	286	52	2,496
2004	2,318	1,780	355	62	1,487
2005	5,571	1,872	174	157	1,855
2006	2,332	1,675	41	280	1,914
2007	3,281	2,197	81	451	2,567
2008	2,346	3,019	29	273	3,263
2009	752	3,366	29	117	3,454
2010	1,209	1,082	170	318	1,230
2011	2,033	348	84	5	269
2012	3,888	794	127	650	1,317
2013	3,168	2,330	429	59	1,960
2014	3,834	3,454	95	155	3,514
2015	4,083	2,435	25	503	2,913
2016	2,642	4,895	30	212	5,077
2017	4,629	3,954	18	182	4,118
2018	4,587	2,309	53	316	2,572
2019	4,549	4,402	139	373	4,636
2020	3,165	3,957	352	438	4,043
2021	2,093	4,081	12	564	4,633
Total	**62,183**	**54,559**	**2,701**	**5,647**	**57,505**
Annual Average	*2,961*	*2,598*	*129*	*269*	*2,738*
Average Past 10 Years	*3,664*	*3,261*	*128*	*345*	*3,478*

Sources: 2020 and 2021 San Francisco Housing Inventory Reports.

Methodology

Examining the characteristic of the parcels that developed and the market conditions prevailing at the time of development can provide a basis for identifying likely sites for future development of multifamily housing from within the larger group of parcels with additional zoned capacity.

In order to identify the characteristics of the parcels and economic conditions that resulted in development of privately financed multifamily units in San Francisco, the project team (the Blue Sky Consulting Group and Planning) developed a database consisting of all the approximately 150,000 parcels in San Francisco, including parcel specific characteristics as well as measures of the housing market conditions and economic circumstances at the time of development.

These data were analyzed using a logistic regression model that estimates the likelihood of development based on several key explanatory variables. Logistic regression is used to model the probability of a discrete, binary outcome (i.e., a parcel develops as multifamily housing or it does not) in which the dependent variable takes on the value of 0 or 1. Explanatory variables include factors that may be correlated with the likelihood that a parcel develops as multifamily housing, including housing prices, construction costs, site specific land use and zoning characteristics, and the "development potential" of individual sites measured as the ratio of potential building size to current size. The model developed offers a comprehensive way to estimate the probability of housing development and the likely number of units on parcels in the City based on both parcel characteristics and current economic trends, addressing requirements that the analysis of non-vacant sites realistically assess housing capacity.

Most Important Factors Contributing to Multifamily Housing Development

While there are many factors that ultimately determine whether a specific parcel develops as housing, empirical analysis and economic theory indicate that the parcel size and "development potential" are important explanatory factors. That is, larger sites were found to be more likely to develop as housing, likely due to the economies of scale for developers in pursuing development projects, with numerous fixed costs for land acquisition and obtaining planning approval, among other factors. In addition, sites that are "under-developed" (i.e., have a high ratio of development potential to current building size) are also more attractive to developers, as these sites tend to generate less in the way of current revenues for property owners relative to the revenue potential associated with residential development for the site.

In addition to these important factors, the prices that developers can charge new residents are also highly important, as the higher the prices, the higher the returns to developers, other things equal. Similarly, lower construction costs are also associated with a higher likelihood of development, as lower costs translate into higher returns for developers, other things equal.

Finally, because it can be difficult to obtain approval for development of sites that have a current residential use or historical designation, variables identifying these sites were included in the model (and found to be statistically significant).

Testing Alternative Specifications

Several alternative models or specifications were tested in the development of the final model, including models that included measures of stock market performance and local unemployment rates (both in San

Francisco and the broader Metropolitan Statistical Area), alternative measures of housing prices and construction costs, and neighborhood designation. In addition, a version of the model was tested using land use designations from the assessor's office; however, these data were found to be both (a) correlated with zoning designations such that including land use in addition to the zoning categories did not add to the explanatory power of the model, and (b) incomplete or missing for many parcels analyzed, resulting in many observations being excluded from the logistic regression.

Data Sources

In order to conduct this analysis, data for each of the more than 150,000 parcels in San Francisco were collected from Planning and other publicly available sources. For each parcel, information was collected regarding the existing land use, zoning, and the potential for future development (i.e., the ratio of allowable building size to current building size). Where factors have changed over time (for example with respect to zoning) data were collected for each year of the study period. To create the development potential variable, a potential building envelop measure was constructed for each parcel in each of the model years. This variable used information about parcel area, setback requirements, density limits, and maximum allowable building height to construct the measure used in the regression model. Finally, the amount of additional development capacity was calculated by dividing the building envelope by the greater of the square footage of the existing building(s) on the parcel for that year or the land area of the parcel if there were no buildings or the information was missing. In addition, information about housing prices and construction costs were included in the model data set for each of the study years. Specifically, the data included in the analysis consisted of the following:[6]

1. Parcel-Specific Data. Data for every parcel in San Francisco were collected for each year of the study period.[7] This information includes attributes that did not change over time such as the parcel's land area and neighborhood, as well as characteristics that may have changed, such as the parcel's zoning designation or maximum allowable building height. Archived annual files for zoning, height and bulk districts, planning districts, special use districts, and land use were used to capture the historical annual data for each parcel and account for any changes over time. In addition, Planning provided information on the maximum allowed density, parking requirements, and setback requirements associated with different planning areas and zoning designations over time. Finally, because parcel identifiers may change over time as parcels are combined or divided, Planning also provided a file that recorded parcel identifier changes over time.

2. Annual Economic Data. Measures of housing prices and construction costs were also collected and integrated to account for economic changes that would have a direct impact on the San Francisco housing market over time, as well as changes in general economic conditions that may influence the amount of housing developed. Housing prices were measured using a San Francisco housing price index published by Zillow, adjusted for inflation using the San Francisco MSA's CPI;

---

6 Note that models including prior land use, economic and demographic data for individual census tracts, and national economic conditions were also tested for inclusion in the regression model; ultimately, these factors were found not to add meaningfully to the explanatory power of the model and were excluded.

7 Note that the unique identifier used in this analysis is the "Map Block Lot Number."

construction costs were measured using the Federal Reserve's real cost index for multifamily residential structures.

3. **Historical Market-Rate Housing Development Data.** Finally, data for market-rate multifamily housing developments completed in San Francisco from 2001 to 2022 were integrated. This list was prepared from Planning's annual Housing Inventory reports. The dataset included the parcel identifier(s) for each project, the year the project was completed, and the number of market-rate and below market-rate (BMR) units for each project.

These data sources were combined to form a single data set for the regression analysis, with one record per year for each of the City's approximately 150,000 parcels over the study period.

Data Overview

Analysis of the resulting database reveals that a relatively small share of the total parcels in San Francisco are suitable for larger scale multifamily development. As shown in Figure 3, more than 85% of the parcels covered by this analysis are zoned for smaller scale residential uses. These parcels zoned RH-1, RH-2 and RH-3 can generally accommodate 1-4 units (including the impact of SB 9, which allows greater density on sites zoned RH-1). Other zoning designations, however, while accounting for a smaller share of parcels, have the capacity to accommodate larger multifamily structures. For example, parcels zoned Office/Commercial can accommodate, on average, just over49 units while parcels with the redevelopment area designation can accommodate on average more than 90 units.

Figure 3: Parcels by Zoning Designation – Modeled Sites Only

Zoning Category	*Parcels*	*% of Parcels*	*Estimated Potential Net Units*	*% of Units*	*Average Net Units per Parcel*
Office / Commercial	1,956	1.3%	96,417	16.8%	49.3
Density Restricted Multifamily	11,357	7.7%	80,592	14.1%	7.1
Form Based Multifamily	5,719	3.9%	77,303	13.5%	13.5
Industrial / Production, Distribution & Repair	1,660	1.1%	0	0.0%	0.0
Public / Open Space	180	0.1%	51,091	8.9%	283.8
Redevelopment Area	39	0.0%	3,603	0.6%	92.4
Residential Single Family (RH-1)	74,673	50.5%	220,590	38.5%	3.0
Residential 2-Family (RH-2, or 2 Units per Lot)	35,157	23.8%	20,844	3.6%	0.6
Residential 3-Family (RH-3) or Res Mixed	17,047	11.5%	21,994	3.8%	1.3
Total	**147,788**	**100.0%**	**572,434**	**100.0%**	**3.9**

Senate Bill 9 and the State Density Bonus

The model results presented below incorporate the likely impact of recent changes to the housing development landscape due to the passage of Senate Bill 9 and changes to the state density bonus contained in Assembly Bill 2345. Specifically, the relationship between development potential and current building size (among other factors, including lot area, construction costs and prices) was used to

estimate the likelihood of development of sites zoned RH1 (single family). In addition, the state density bonus was applied to eligible parcels to estimate an effective increase in development potential for those sites, which resulted in an increase in the probability of development and expected units developed. Specifically, for parcels that had a base zoned capacity for more than 10 units, the maximum building envelope was increased by 40 percent (rather than the statutory maximum of 50 percent) above the currently zoned maximum. Further, eligible parcels were assumed to use the density bonus 60 percent of the time, based on the fact that in recent years approximately 60 percent of projects have chosen to use the density bonus.

Results

The regression-based model provides an estimate of the total number of units that would be expected to be developed over the eight-year RHNA period given the characteristics of each parcel and broader economic trends and conditions. For each included parcel, a probability of development was estimated and multiplied by the potential number of units that could be constructed at that site to arrive at an estimated number of units.

The model estimates the probability of development based on a series of parcel specific characteristics as well as city-wide measures of housing prices and construction costs. As shown in Figure 4, the included variables were highly statistically significant.[8]

While the model provides a parcel-level estimate of units to be produced, the results are best interpreted in aggregate. Planning has used the model results to estimate that 9,186 units are probable through privately funded multifamily housing development over the RHNA period on parcels available for residential development in the city and not already accounted for in the residential development pipeline or included in a development agreement.

Interpreting the Regression Coefficients

By basing the estimates of likely future development on historical observations of actual development projects, the model developed by the Blue Sky Consulting Group offers a more realistic approach to estimating capacity for RHNA than has been used in the past.

For each parcel in the City where housing is allowed, the model estimates the likely number of units based on the regression results, calculated as the probability of development for the site multiplied times the number of units allowed on that site.

8 Statistical significance for the Logit model is indicated by the value in the column "Prob>ChiSq." A small value in this column indicates that the result is very unlikely to be due to random chance. All of the variables, with the exception of the "Zoning = Public/Open Space" variable, were significant at the 95% confidence level or above.

Figure 4. Logistic Regression Analysis Results

Explanatory Variables – Descriptions	*Values for Selected Model*	
	coeff	*Prob>Ch Sq*
Intercept	(10.2835)	0.0000
Parcel has Historic Status (Dummy Variable)	(0.5213)	0.0000
Parcel has Existing Residential Use (Dummy Variable)	(1.1345)	0.0000
SF Housing Price Index (Zillow), Real	0.0511	0.0000
Federal Reserve Multifamily Housing Index, Real	(0.0391)	0.0000
Potential Building Envelope in 1000 sq ft	0.0007	0.0199
Potential Building Envelope / Existing sq ft	0.0763	0.0000
Zoning Dummy Variables:		
Zoning = Office/Commercial	3.2714	0.0000
Zoning = Density Restricted Multifamily	2.7671	0.0000
Zoning = Form Based Multifamily	3.6281	0.0000
Zoning = Industrial / Production, Distribution & Repair	2.2291	0.0000
Zoning = Public/Open Space	(1.4265)	0.1561
Zoning = Redevelopment Area	3.6509	0.0000
Zoning = Residential 2-Family (2 Units per Lot)	1.3510	0.0000
Zoning = Residential 3-Family or Residential Mixed-1 (1/800 sqft)	1.4429	0.0000

Note: Omitted zoning variable is RH1 (Residential Single Family); coefficients shaded in yellow are statistically significant at the 95 percent level.

Larger, positive coefficient values (as presented in Figure 4) indicate that the variable is associated with a higher likelihood of development while smaller or negative values are associated with a lower likelihood of development. These model estimates align with intuitive expectations and economic theory. For example, larger sites with no existing structures or small existing structures and where greater numbers of housing units are allowed (as measured by the "Potential Building Envelope/Existing sq ft" variable) are likely to have more estimated units in the model (coefficient of 0.0763). Conversely, parcels with an existing residential use ("Parcel Has Existing Residential Use") are less likely to be a site for future development, as demonstrated by its negative coefficient value (coefficient value of -1.1345).

Housing prices (a key component of developer return) were measured through the "SF Housing Price Index" variable, based on data collected by Zillow on the prices for multifamily housing in San Francisco. The coefficient of 0.0511 indicates that higher prices are associated with an increase in the likelihood of development. Construction costs were measured with the inclusion of a construction cost index ("Federal Reserve Multi Family Housing Index"). The regression coefficient of -0.0391 indicates that higher construction costs are associated with a lower likelihood of development.

Among the various zoning designations, parcels currently zoned "Form Based Multifamily" and "Redevelopment Area" are the most likely to be developed as multi-family residential, all else equal; parcels zoned for open space or small residential are the least likely to be developed as multifamily housing.

Figure 5 presents the expected number of units to be produced by zoning category. As shown in Figure 5, parcels zoned as "Form Based Multifamily" (i.e., residential zoning with height restrictions and setback requirements but no specific density limits) are anticipated to produce 4,223 units over the 8-year RHNA period. The much larger number of sites zoned RH1, RH2, or RH3, in contrast, are expected to generate just over 550 net units over this period.

Figure 5. **Estimated Units by Zoning Designation**

Zoning Category	*Parcels*	*Forecast Net Units*	*Percent*
Office / Commercial	1,956	2,845	31.0%
Density Restricted Multifamily	11,357	1,389	15.1%
Form Based Multifamily	5,719	4,223	46.0%
Industrial / Production, Distribution & Repair	1,660	0	0.0%
Public / Open Space	180	58	0.6%
Redevelopment Area	18	114	1.2%
Residential Single Family (RH-1)	74,673	454	4.9%
Residential 2-Family (RH-2, or 2 Units per Lot)	35,157	40	0.4%
Residential 3-Family (RH-3) or Res Mixed	17,047	63	0.7%
Total	**147,768**	**9,186**	**100.0%**

Distribution of Development Probabilities

According to the model results, most parcels in San Francisco have a low likelihood of development as multifamily housing; given that there are approximately 150,000 parcels in the city, but just a handful of multifamily residential projects each year, this is the expected result. Nevertheless, while the vast majority of parcels will not be developed as multifamily housing in a given year (and will produce zero new units), some parcels will develop each year and will produce more than their probability-adjusted "expected number" of units. Therefore, by aggregating the results across parcels, a realistic estimate of the total number of units expected to develop over the study period can be estimated. Figure 6 presents data regarding the distribution of the probability of development. As shown in Figure 6, the probability of development over the 8-year period across all parcels is just 0.40%. The probability varies by zoned capacity for the parcel; parcels with zoned capacity below 10 units average just 0.32% probability of development over the 8-year RHNA period, while those that can accommodate 10 to 50 units or more than 50 units averaging 2.12% and 2.42% probability of development, respectively. Overall, 90% of all parcels have less than a one percent probability of development.

Figure 6. Probability of Development by Zoned Capacity

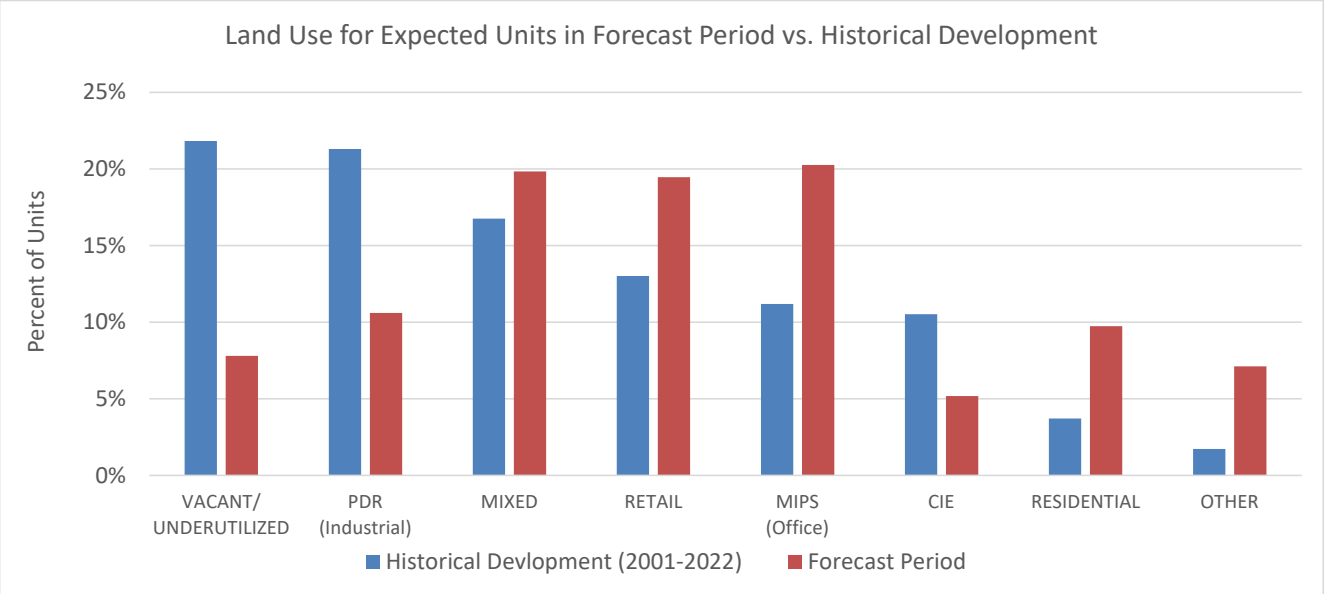
Percentiles	*All Parcels in Model*	*By Zoned Capacity*		
		< 10 Units	*10 – 50 Units*	*> 50 Units*
1st	0.10%	0.10%	0.13%	0.02%
5th	0.11%	0.11%	0.35%	0.03%
10th	0.16%	0.16%	0.48%	0.12%
25th	0.17%	0.17%	0.81%	0.91%
50th	0.20%	0.20%	1.53%	1.93%
75th	0.21%	0.21%	2.79%	3.17%
90th	0.73%	0.64%	4.63%	4.96%
95th	1.68%	1.03%	6.31%	5.75%
99th	3.84%	2.34%	7.73%	8.25%
Mean	**0.40%**	**0.32%**	**2.12%**	**2.42%**
Parcels	**147,768**	**141,237**	**5,131**	**1,400**

Analysis of Prior Land Uses

Figure 7, below, shows the prior land use associated with development that occurred during the study period as well as the expected development during the 8-year RHNA period according to the regression model results. Previously underutilized or vacant sites were the most common type of prior land use historically at around 22% of the sites where multifamily housing was developed and are expected to account for approximately 8% of the units over the 8 year RHNA period.[9] In addition to previously vacant/underutilized sites, residential development occurred on sites that had a previous industrial use (denoted as Production, Distribution and Repair or PDR). Other common prior uses include mixed use, retail, and office (designated as "MIPS" or Management, Information or Professional Services). New multifamily development also occurred (to a lesser extent) on sites that had a previous residential use or were designated cultural, institutional or educational (CIE); model results indicate that some development will likely occur on such sites going forward, though to a lesser extent than on sites with other existing uses such as office (MIPS), retail or mixed use.

9 The results presented in Figure 7 exclude housing developed pursuant to a development agreement or on parcels designated as redevelopment areas.

Figure 7. Land Use Prior to Development



Requirements in Government Code Section 65583.2

Government Code Section 65583.2 imposes certain requirements on a local government's inclusion of sites designated as suitable for residential development. Specifically, Section 65583.2 requires that for designated sites "the city or county shall specify the additional development potential for each site within the planning period and shall provide an explanation of the methodology used to determine the development potential. The methodology shall consider factors including the extent to which existing uses may constitute an impediment to additional residential development, the city's or county's past experience with converting existing uses to higher density residential development, the current market demand for the existing use, an analysis of any existing leases or other contracts that would perpetuate the existing use or prevent redevelopment of the site for additional residential development, development trends, market conditions, and regulatory or other incentives or standards to encourage additional residential development on these sites."

Each of these factors has been addressed by the current methodology, as identified below:

1. "the additional development potential for each site" has been addressed through the inclusion in the regression model of the calculated zoned capacity for each site relative to the size of the current structure on each site. In fact, this ratio (the development potential) is the key instrument variable included in the model and is highly statistically significant.
2. "the extent to which existing uses may constitute an impediment to additional residential development" and "the city's or county's past experience with converting existing uses to higher density residential development" has been addressed by an analysis of the land use existing on sites that developed as privately financed multi-family housing during the study period (see

"Analysis of Prior Land Uses" on page 12). In addition, prior land use in terms of residential use and historically protected sites was included in the regression analysis through the residential and historical designation variables. See Appendix 3 for a series of case studies documenting residential housing development on previously non-vacant sites in the City.

3. "the current market demand for the existing use" is addressed through the inclusion of the price variable in the regression model, which is correlated with market demand and the potential return for developers.

4. "an analysis of any existing leases or other contracts that would perpetuate the existing use or prevent redevelopment of the site for additional residential development" is addressed by an analysis of lease duration, which shows that the duration of leases did not change significantly during the study period.[10]

5. "development trends, market conditions, and regulatory or other incentives or standards to encourage additional residential development" is addressed through the regression model by inclusion of construction cost and market price variables as well as the adjustments for SB 9 and the state density bonus.

10 Although limited data are available, see CBRE, "How does the economic cycle influence the length of office leases?" which found that "Generally, lease term lengths have been quite stable over the past 35 years."